Methods

mRNA was isolated from the mammary epithelial cells from 5 replicates with either wild type (wt), *Gain Myc-Pvt1*, *Gain Myc* or *Gain Pvt1* merged by genotype and barcoded cDNA was generated which was sequenced on the Illumina Hi-SEQ platform as 2*50 paired end reads at an average depth of ~20 million reads per library. Data was mapped to the MM10 genome using Bowtie short read mapper[1] via TOPHAT mapping algorithm. Cuffdiff was used with the igenomes MM10 UCSC .gtf file (downloaded 5-23-2014) to identify differentially expressed transcripts between samples and summarize the data into both FPKM values and normalized read counts[2]. In addition to calling differential expressed genes by cuffdiff, we further cleaned up the dataset by 1) identifying all genes that were present with greater than 20 read (or normalized) counts across the full dataset and 2) rounding up all values below 0.1 FPKM up to 0.1 in order to minimize the effect of dividing by essentially meaningless FPKMS close to zero.

Breast Cancer RNA-SEQ data was downloaded from the TCGA (https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp).

References

1. Langmead, B. & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9.
2. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L. & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-78.